CLAIMS

1. A method comprising:

extracting an intensity feature, a timbre feature, and a rhythm feature from a music clip;

classifying the music clip into a mood group based on the intensity feature; and

classifying the music clip into an exact music mood from the mood group based on the timbre feature and the rhythm feature.

2. A method as recited in claim 1, wherein the extracting comprises:

converting the music clip into a uniform music clip having a uniform format;

dividing the uniform music clip into a plurality of frames; and

dividing each frame into a plurality of octave-based frequency sub-bands.

3. A method as recited in claim 2, wherein the extracting an intensity feature comprises:

calculating a root mean-square (RMS) signal amplitude for each sub-band of each frame;

summing the RMS signal amplitudes across the sub-bands of each frame to determine a frame intensity for each frame; and

averaging the frame intensities to determine the intensity feature for the music clip.

4. A method as recited in claim 2, wherein the extracting a timbre feature comprises:

calculating spectral shape features for each frame;

calculating spectral contrast features for each frame; and

representing the timbre feature with one or more of the spectral shape features and/or the spectral contrast features.

5. A method as recited in claim 2, wherein the extracting a rhythm feature comprises:

extracting an amplitude envelope from the lowest sub-band and the highest sub-band of each frame across the uniform music clip;

estimating a difference curve of the amplitude envelope; and

detecting peaks above a threshold within the difference curve, the peaks being instrumental onsets.

6. A method as recited in claim 5, wherein the extracting a rhythm feature further comprises:

extracting an average rhythm strength of the instrumental onsets;

extracting a rhythm regularity value based on the average of the maximum three peaks in the difference curve; and

extracting a rhythm tempo based on a common divisor of peaks in the difference curve.

7. A method as recited in claim 1, wherein the classifying the music clip into a mood group comprises:

determining the probability of a first mood group based on the intensity feature;

determining the probability of a second mood group based on the intensity feature;

selecting the first mood group if the probability of the first mood group is greater than or equal to the probability of the second mood group; and

otherwise selecting the second mood group.

8. A method as recited in claim 1, wherein the classifying the music clip into a mood group comprises classifying the music clip into a mood group selected from the group comprising:

a contentment and depression mood group; and

an exuberance and anxious mood group.

9. A method as recited in claim 1, wherein the mood group includes a first mood and a second mood, the classifying the music clip into an exact music mood comprising:

determining the probability of the first mood based on the timbre feature and the rhythm feature;

determining the probability of the second mood based on the timbre feature and the rhythm feature;

selecting the first mood as the exact mood if the probability of the first mood is greater than or equal to the probability of the second mood; and

otherwise selecting the second mood as the exact mood.

10. A method as recited in claim 9, wherein the mood group is selected from the group comprising:

a first mood group that includes a contentment mood and a depression mood; and

a second mood group that includes an exuberance mood and an anxious mood.

11. A processor-readable medium comprising processor-executable instructions configured for:

extracting features from a music clip;

selecting a first mood group or a second mood group based on a first feature; and

determining an exact mood from within the selected mood group based on a second feature and a third feature.

12. A processor-readable medium as recited in claim 11, wherein the extracting comprises:

down-sampling the music clip into a uniform format;

dividing the music clip into a plurality of frames; and

dividing each frame into a plurality of frequency sub-bands.

13. A processor-readable medium as recited in claim 12, wherein the down-sampling comprises converting the music clip into a 16 KHz, 16 bit, mono-channel uniform sample.

14. A processor-readable medium as recited in claim 12, wherein the dividing the music clip into a plurality of frames comprises dividing the music clip into non-overlapping, 32 microsecond-long frames.

15. A processor-readable medium as recited in claim 12, wherein the dividing each frame into a plurality of frequency sub-bands comprises dividing each frame into seven frequency sub-bands, each sub-band being an octave sub-band.

16. A processor-readable medium as recited in claim 12, wherein the extracting comprises extracting an intensity feature.

17. A processor-readable medium as recited in claim 16, wherein the extracting an intensity feature comprises extracting an intensity feature for each frame, the processor-readable medium comprising further processor-executable instructions configured for calculating a root mean-square (RMS) signal amplitude for each sub-band of each frame.

18. A processor-readable medium as recited in claim 17, comprising further processor-executable instructions configured for summing the RMS signal amplitudes across the sub-bands of each frame to determine a frame intensity feature for each frame.

19. A processor-readable medium as recited in claim 18, comprising further processor-executable instructions configured for averaging the frame intensity features across all frames to determine a music clip intensity feature.

20. A processor-readable medium as recited in claim 12, wherein the extracting comprises extracting a timbre feature.

21. A processor-readable medium as recited in claim 20, wherein the extracting a timbre feature comprises extracting a timbre feature for each frame, and wherein the extracting a timbre feature for each frame comprises:

determining spectral shape features;

determining spectral contrast features; and

representing the timbre feature with the spectral shape features and the spectral contrast features.

22. A processor-readable medium as recited in claim 21, wherein the determining spectral shape features comprises determining one or more shape features from the group comprising:

a frequency centroid of a frame;

a frequency bandwidth of a frame;

a frequency roll off of a frame; and

a spectral flux of a frame.

23. A processor-readable medium as recited in claim 21, wherein the determining spectral contrast features comprises determining one or more contrast features from the group comprising:

a spectral peak in a sub-band of a frame;

a spectral valley in a sub-band of a frame; and

a spectral average of all spectral components in a sub-band of a frame.

24. A processor-readable medium as recited in claim 12, wherein the extracting comprises extracting a rhythm feature.

25. A processor-readable medium as recited in claim 24, wherein the extracting a rhythm feature comprises:

extracting an amplitude envelope from a lowest sub-band and a highest sub-band;

estimating a difference curve of the amplitude envelope; and

detecting peaks above a threshold within the difference cure, the peaks being bass instrumental onsets.

26. A processor-readable medium as recited in claim 25, wherein the extracting a rhythm feature further comprises:

extracting an average rhythm strength of the instrumental onsets;

extracting a rhythm regularity value based on an average of the maximum three peaks in the difference curve; and

extracting a rhythm tempo based on a common divisor of peaks in the difference curve.

27. A processor-readable medium as recited in claim 11, wherein the selecting comprises:

determining the probability of the first mood group given the first feature;

determining the probability of a second mood group given the first feature;

selecting the first mood group if the probability of the first mood group is greater than or equal to the probability of the second mood group; and

otherwise selecting the second mood group.

28. A processor-readable medium as recited in claim 27, wherein the first feature is an intensity feature.

29. A processor-readable medium as recited in claim 27, wherein the first mood group comprises a contentment mood and a depression mood, and the second mood group comprises an exuberance mood and an anxious mood.

30. A processor-readable medium as recited in claim 11, wherein the selected mood group comprises a first mood and a second mood, and the determining an exact mood from within the selected mood group comprises:

determining the probability of the first mood given the second and third features;

determining the probability of a second mood given the second and third features;

selecting the first mood as the exact mood if the probability of the first mood is greater than or equal to the probability of the second mood; and

otherwise selecting the second mood as the exact mood.

31. A processor-readable medium as recited in claim 30, wherein the determining the probability of the first mood given the second and third features comprises:

determining a weighted first probability, the weighted first probability being a first weight multiplied by the probability of the first mood based on the second feature;

determining a weighted second probability, the weighted second probability being a second weight multiplied by the probability of the first mood based on the third feature, wherein the sum of the first weight and the second weight is equal to one; and

summing the weighted first probability and the weighted second probability.

32. A processor-readable medium as recited in claim 30, wherein the determining the probability of the second mood given the second and third features comprises:

determining a weighted first probability, the weighted first probability being a first weight multiplied by the probability of the second mood based on the second feature;

determining a weighted second probability, the weighted second probability being a second weight multiplied by the probability of the second mood based on the third feature, wherein the sum of the first weight and the second weight is equal to one; and

summing the weighted first probability and the weighted second probability.

33. A processor-readable medium as recited in claim 30, wherein the second feature is a timbre feature and the third feature is a rhythm feature.

34. A processor-readable medium as recited in claim 11, wherein the extracting comprises:

extracting an intensity feature;

extracting a timbre feature; and

extracting a rhythm feature.

35. A processor-readable medium as recited in claim 11, comprising further processor-executable instructions configured for:

constructing a Gaussian Mixture Model (GMM) to model each feature; and

estimating parameters of a Gaussian component and mixture weights within the GMM using an Expectation Maximization (EM) algorithm.

36. A processor-readable medium as recited in claim 35, comprising further processor-executable instructions configured for initializing the GMM using a K-means algorithm.

37. A computer comprising:

a music clip; and

a mood detection algorithm configured to classify the music clip as a music mood according to music features extracted from the music clip.

38. A computer as recited in claim 37, further comprising a music feature extraction tool configured to extract the music features.

39. A computer as recited in claim 38, further comprising a hierarchical music mood detection process configured to determine a mood group based on a first music feature and an exact music mood from within the mood group based on a second and third music feature.

40. A system comprising:

a music clip;

a feature extraction tool configured to extract music features from the music clip; and

a hierarchical music mood detection module configured to classify the music clip into a music mood based on the music features.